CLAIMS

What is claimed is:

1.   A processor-based method, comprising:

generating a plurality of different functions that correlate datapoints of a dataset through a regression clustering algorithm; and

determining directives by which to categorize new data into the dataset through a classification algorithm.

2.   The processor-based method of claim 1, further comprising receiving a set of new data, wherein the set of new data comprises a known value for a first variable parameter of the dataset and is absent of a value for a second variable parameter of the dataset.

3.   The processor-based method of claim 2, further comprising:

selecting a function having the highest probability of representing the correlative relationship between the first and second variable parameters; and

predicting the value of the second variable parameter from the selected function.

4.   The processor-based method of claim 2, further comprising predicting a value with the highest probability of correlating to a plurality of the generated functions for the second variable parameter.

5.   The processor-based method of claim 2, further comprising predicting a plurality of possible values for the second variable parameter based upon the value of the first variable parameter.

6.   The processor-based method of claim 1, wherein said generating the plurality of different functions comprises applying a regression algorithm and a K-Harmonic Means performance function on the datasets.

7.   The processor-based method of claim 1, wherein said generating the plurality of different functions comprises applying a regression algorithm and a K-Means performance function on the datasets.

8.   The processor-based method of claim 1, wherein said generating the plurality of different functions comprises applying a regression algorithm and an Expectation Maximization performance function on the datasets.

9.   A storage medium comprising program instructions executable by a processor for:

    generating a plurality of different functions that correlate variable parameters of a dataset;

    based upon the different functions, determining classification directives by which to group new data into the dataset; and

    based upon the classification directives, selecting one of the functions to predict a value of a variable parameter associated with one or more known input values of the new data.

10.   The storage medium of claim 9, wherein the program instructions for selecting comprise program instructions for selecting the function with the highest probability of representing a correlative relationship between the one or more known input values and the variable parameter.

11.   The storage medium of claim 9, wherein the program instructions for selecting the function with the highest probability comprise program instructions for:

    determining a kernel density estimation for each the plurality of functions;

    summing the kernel density estimations; and

    determining, from the kernel density estimation summation, the probability of each function representing a correlative relationship between the value of the variable parameter and the one or more known input values.

12.   The storage medium of claim 9, wherein the variable parameter associated with the predicted value is one of the variable parameters correlated by the plurality of different functions.

13.   The storage medium of claim 9, wherein the variable parameter associated with the predicted value is unrelated to the variable parameters correlated by the plurality of different functions.

14.   The storage medium of claim 9, wherein the program instructions for generating comprise program instructions for regressively clustering the dataset.

15.   The storage medium of claim 9, wherein the program instructions for generating comprise program instructions for generating a plurality of functions for a partially labeled dataset.

16.   The storage medium of claim 9, wherein the program instructions for generating comprise program instructions for generating a plurality of functions from a partially completed dataset.

17.   A storage medium comprising program instructions executable by a processor for:

    generating a plurality of different functions that correlate variable parameters of a dataset;

    receiving new data associated with the dataset, wherein the new data comprises known input values for a first set of the variable parameters and is substantially absent of values for a second set of the variable parameters; and

    determining a plurality of possible values for at least one value of the second set of the variable parameters based upon the known input values of the first set of variable parameters and the plurality of different functions.

18.   The storage medium of claim 17, wherein the program instructions for determining comprise program instructions for determining the same number of possible values as the number of generated functions.

19.   The storage medium of claim 17, further comprising program instructions executable by the processor for determining classification directives by which to group new data into the dataset based upon the different functions.

20.   The storage medium of claim 17, wherein the program instructions for generating comprise program instructions for:

      selecting a set number of functions correlating variable parameters of a dataset;

      determining distances between datapoints of the dataset and values correlated with the set number of functions;

      calculating harmonic averages of the distances;

      regressing the set number of functions using datapoint probability and weighting factors associated with the determined distances;

      repeating said determining and calculating for the regressed set of functions;

      computing a change in harmonic averages for the set number of functions prior to and subsequent to said regressing; and

      reiterating said regressing, repeating and computing upon determining the change in harmonic averages is greater than a predetermined value.

21.   A system, comprising:

      a dataset;

      a means for generating a plurality of different functions that correlate datapoints of the dataset; and

      a means for classifying new data into the dataset with respect to the plurality of different functions.

22. The system of claim 21, further comprising a means for predicting one or more values for a variable parameter associated with input values of the new data.

23. The system of claim 22, wherein the means for predicting the one or more values comprises a means for determining a value of the variable parameter for each of the plurality of functions.

24. The system of claim 22, wherein the means for predicting the one or more values comprises:

a means for determining the function with the highest probability of representing the correlative relationship between the variable parameter and input values; and

a means for determining the value of the variable parameter from the function.

25. The system of claim 22, wherein the means for predicting the one or more values comprises a means for determining a value with the highest probability of correlating to a plurality of the different functions for the variable parameter.

26. A system, comprising:

an input port configured to access a dataset; and

a processor configured to:

regressively cluster the dataset to generate functions that correlate datapoints of the dataset; and

determine classifiers by which to correlate new data with respect to the generated functions.

27. The system of claim 26, wherein the processor is further configured to predict values of one or more variable parameters associated with the new data.

28. The system of claim 27, wherein the processor is configured to:

select a function which best represents the new data; and

predict the values of the one or more variable parameters from the selected function.

29. The system of claim 27, wherein the processor is configured to predict a value with the highest probability of correlating to a plurality of the different functions for the variable parameter.

30. The system of claim 27, wherein the processor is configured to predict a plurality of possible values for the one or more variable parameters.